

<https://helda.helsinki.fi>

Rational coordination with no communication or conventions

Goranko, Valentin

2020-07-11

Goranko , V , Kuusisto , A & Rönholm , R 2020 , ' Rational coordination with no communication or conventions ' , Journal of Logic and Computation , vol. 30 , no. 6 , pp. 1183-1211 . <https://doi.org/10.1093/logcom/exaa032>

<http://hdl.handle.net/10138/332355>

<https://doi.org/10.1093/logcom/exaa032>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Rational coordination with no communication or conventions

Valentin Goranko¹, Antti Kuusisto^{2,3}, Raine Rönholm^{3,4}

¹Stockholm University, Sweden

²University of Helsinki, Finland

³Tampere University, Finland

⁴ENS Paris-Saclay, France

Abstract

We study pure coordination games where in every outcome, all players have identical payoffs, ‘win’ or ‘lose’. We identify and discuss a range of ‘purely rational principles’ guiding the reasoning of rational players in such games and compare the classes of coordination games that can be solved by such players with no preplay communication or conventions. We observe that it is highly nontrivial to delineate a boundary between purely rational principles and other decision methods, such as conventions, for solving such coordination games.

1 Introduction

Coordination games ([16]) are games in strategic form with several pure strategy Nash equilibria with the same or comparable payoffs for every player. In these games, all players have the mutual interest to select one of these equilibria. In *pure coordination games* ([16]), aka *games of common payoffs* ([17]), all players in the game receive the same payoffs and thus the players have fully aligned preferences to coordinate with each other in order to reach the best possible outcome for everyone. In this paper we study one-step *pure win-lose coordination games* (WLC games) in which all payoffs are either 1 (i.e., *win*) or 0 (i.e., *lose*).

Clearly, if players can communicate when playing a pure coordination game with at least one winning outcome, then they can simply agree on a winning strategy profile, so the game is trivialised. What makes such games non-trivial is limited (or non-existing)

possibility of preplay communication amongst the players¹, meaning that the players must make their choices based on individual reasoning—without any contact with the other players before (or during) playing the game.

There are many natural real-life situations where such coordination scenarios occur. We give two examples:

- (A) two cars driving towards each other (on a narrow street) that can avoid a collision by swerving either to the right or left,
- (B) a group of n people who get separated in a city and must individually decide on a place where to get together (‘regroup’), supposing the group members do not have any way of contacting each other.

The notion of *convention* is an important concept that emerges in the context of coordination. Following Lewis’ seminal book [16], much of the literature on the topic focuses on *social conventions* which are not explicit agreements established via communication, but rather regularities emerging in social behaviour that every individual believes everyone to follow. If indeed followed by everyone, such conventions help resolve social coordination problems. For example, most words used in natural languages are examples of naturally emerged social conventions, whereas using the metric system instead of some other set of measures is likely to be mainly based on an explicit agreement.

Both types of conventions may be suitable in different situations. For instance, in the scenario (A) above, a collision could be avoided by using the explicitly agreed convention (traffic rule) that cars should always swerve to the right, whereas in (B), everyone could go to a famous meeting spot in the city established by a social convention, e.g., the main entrance of the main railway station. The main entrance of the main railway station is an example of a possible *focal point* (Schelling [25]) that sticks out and therefore may be likely to emerge as, for example, an obvious meeting point.

In most of this paper we assume that players share *no conventions* at all, and we also assume *no preplay communication*. Thus, in our basic setting, the players play completely independently of each other and could be assumed to come from entirely different kinds of cultures, or even from different galaxies for that matter. The principal question in this paper is the following:

What kinds of reasoning can be accepted as purely rational, based on no communication and no conventions? And which pure win-lose coordination games can be solved by such reasoning?

Thus we identify “*purely rational principles*” that *every* ideally rational player ought to follow in *every* WLC game. We also study the hierarchy of games solvable by such principles. For instance, it is intuitively clear that coordination by pure rationality is not

¹Note that, while the common use of ‘preplay communication’ in game theory means communication before the given game is played, here we also mean communication before the players are *even presented with the game*.

possible in the example situations (A) and (B) above. However, we will see that there are many natural coordination scenarios in which it seems clear that rational players can coordinate successfully by following purely rational reasoning principles, without preplay communication or conventions.

Initially, we assume only *individual rationality*, i.e., that every player acts rationally with the aim to win the game but without assumptions about the other players' rationality. Later we additionally assume *common belief in rationality*, i.e., that every player is individually rational and that it is commonly believed amongst the players that every player is rational.

Towards the end of the article, we move from pure rationality to the setting with conventions. Thus, as conventions could indeed arise from preplay communication, we now move to the setting where the assumption of “*no preplay communication*” is essentially banished.² We consider a special class of conventions that we call *structural conventions*. These are conventions only based on structural properties of the game (i.e., essentially properties invariant under game isomorphisms) rather than on ad hoc features, such as names of the choices in the game.

All through the paper we make the following assumptions.

- We only consider complete information games, i.e., the game structure is common knowledge amongst the players.
- It is common knowledge amongst the players that they all have the same goal, which is to win, i.e., to select together a winning choice profile.

The main outcomes of our study are as follows:

- (i) Concerning the scenario with no communication and no conventions, we identify several different kinds of reasoning principles and provide justifications for them. These justifications have varying levels of common acceptability. We observe that the boundary of purely rational principles and other principles is highly nontrivial to demarcate³. Indeed, the question of what constitutes a purely rational principle seems open-ended and depends on different background assumptions.
- (ii) On the other hand, the class of WLC games that are solvable by using structural conventions is precisely characterized in Section 5 in terms of structural properties (involving symmetries) of games.

There exist a wide range of works closely related to the current paper. Indeed, coordination and rationality are natural and relevant topics and have thus been studied in various different contexts, e.g., in [4], [8], [9], [10], [27]. The theory of focal points has been extensively developed in the context of coordination, e.g., in [25], [10], [27], [20], [28], [15], [5], [2]. Conventions have also been studied, e.g., in [16], [26], [10]. Furthermore, we note the close conceptual relationship of the present study with the notion of *rationalisability* of

²In particular, we are mainly interested in the scenario where the players can adopt conventions (for example through negotiations) *only before* being presented with the particular game to play.

³Schelling shares this view on pure coordination games (see [25], p. 283, n. 16).

strategies [3], [22], [7], which is particularly important in epistemic game theory. We also mention two recent relevant works—related to logic—to which the observations and results in the present paper could be directly applied: in [14], two-player coordination games are related to a variant of *Coalition Logic*⁴, and in [1], coordination scenarios are analysed with respect to the game-theoretic semantics of *Independence Friendly Logic*. Finally, the study on structural conventions with respect to ‘random game graphs’ in [19] adopts the definitions that were originally conceived in the work leading to the current paper.

This paper is a substantially extended version of the conference paper [12], while [13] is an extended abstract of a workshop presentation of a part of this work.

2 Pure win-lose coordination games

2.1 The setting

A *pure win-lose coordination game* G is a strategic form game with n players $(1, \dots, n)$ whose available *choices* (*moves*, *actions*) are given by sets $\{C_i\}_{i \leq n}$. The set of winning *choice profiles* is presented by an n -ary *winning relation* W_G . For technical convenience and simplification of some definitions, we present these games as *relational structures* (see, e.g., [6]). A formal definition follows.

Definition 2.1 *An n -player **win-lose coordination game** (WLC game) is a relational structure $G = (A, C_1, \dots, C_n, W_G)$ where A is a finite domain of **choices**, each $C_i \neq \emptyset$ is a unary predicate (i.e., a subset of A), representing the choices of player i , s.t. $C_1 \cup \dots \cup C_n = A$, and W_G is an n -ary relation in A such that $W_G \subseteq C_1 \times \dots \times C_n$. For technical convenience, here we also assume that the players have pairwise disjoint choice sets, i.e., $C_i \cap C_j = \emptyset$ for every $i, j \leq n$ such that $i \neq j$. A tuple $\sigma \in C_1 \times \dots \times C_n$ is called a **choice profile** for G and the choice profiles in W_G are called **winning choice profiles**.*

We use the following terminology for any WLC game $G = (A, C_1, \dots, C_n, W_G)$.

- The **losing relation** of G is the relation $L_G := C_1 \times \dots \times C_n \setminus W_G$. A choice profile $\sigma \in L_G$ is called **losing**.
- The **complementary game** of G is the game $\overline{G} := (A, C_1, \dots, C_n, L_G)$.
- Let $A_i \subseteq C_i$ for every $i \leq n$. The **restriction** of G to (A_1, \dots, A_n) is the game $G \upharpoonright (A_1, \dots, A_n) := (A_1 \cup \dots \cup A_n, A_1, \dots, A_n, W_G \upharpoonright A_1 \times \dots \times A_n)$.
- For every choice $c \in C_i$ of a player i , the **winning extension** of c in G is the set $W_G^i(c)$ of all tuples $\tau \in C_1 \times \dots \times C_{i-1} \times C_{i+1} \times \dots \times C_n$ such that the choice profile obtained from τ by adding c to the i -th position is winning. We define the **losing extension** of c in G analogously.

⁴In fact, the initial motivation for the present work came from concerns with the semantics of Alternating time temporal logic ATL, extending Coalition Logic.

- A choice $c \in C_i$ of a player i is **(surely) winning**, respectively **(surely) losing**, if it is guaranteed to produce a winning (respectively losing) choice profile regardless of what choices the other player(s) make. Note that c is a winning choice iff $W_G^i(c) = C_1 \times \cdots \times C_{i-1} \times C_{i+1} \times \cdots \times C_n$. Similarly, c is a losing choice iff $W_G^i(c) = \emptyset$.

Example 2.2 *We present here a 3-player coordination story which will be used as a running example hereafter. Three robbers, Casper, Jesper and Jonathan⁵ are planning to quickly steal a cake from the bakery of Cardamom Town while the baker is out. They have two possible plans to enter the bakery: either (a) to break in through the front door or (b) to sneak in through a dark open basement. For (a) they need a crowbar and for (b) a lantern. The baker keeps the cake on top of a high cupboard, and the robbers can only reach it by using a ladder.*

When approaching the bakery, Casper is carrying a crowbar, Jesper is carrying a ladder and Jonathan is carrying a lantern. However, the robbers cannot agree whether they should follow plan (a) or plan (b). While the robbers are quarreling, suddenly Constable Bastian appears and the robbers all flee to different directions. After this the robbers have to individually decide whether to go to the front door (by plan (a)) or to the basement entrance (by plan (b)). They must take the right decision fast before the baker returns.

The scenario described here can naturally be modeled as a WLC game. We relate Casper, Jesper and Jonathan with players 1, 2 and 3, respectively. Each player i has two choices a_i and b_i which correspond to either going to the front door or to the basement entrance, respectively. The robbers succeed in obtaining the cake if both Casper and Jesper go to the front door (whence it does not matter what Jonathan does). Or, alternatively, they succeed if both Jonathan and Jesper go to the basement (whence the choice of Casper is irrelevant). This coordination scenario corresponds to the following WLC game $G^ = (\{a_1, b_1, a_2, b_2, a_3, b_3\}, C_1, C_2, C_3, W_{G^*})$, where for each player i , $C_i = \{a_i, b_i\}$ and $W_{G^*} = \{(a_1, a_2, a_3), (a_1, a_2, b_3), (a_1, b_2, b_3), (b_1, b_2, b_3)\}$. (For a graphical presentation of this game, see Example 2.3 below.)*

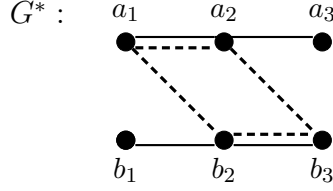
2.2 Presenting WLC games as hypergraphs

The n -ary winning relation W_G of an n -player WLC game G defines a *hypergraph* on the set of all choices. We give visual presentations of hypergraphs corresponding to WLC games as follows. The choices of each player are displayed as columns of nodes starting from the choices of player 1 on the left and ending with the column with choices of player n . The winning relation consists of lines that go through some choice of each player⁶. This kind of graphical presentation of a WLC game G will be called a *game graph (drawing)* of G . (Note that game graphs of 2-player WLC games are simply bipartite graphs.)

⁵This example is based on the children's book *When the Robbers Came to Cardamom Town* by Thorbjørn Egner, featuring the characters Casper, Jesper and Jonathan.

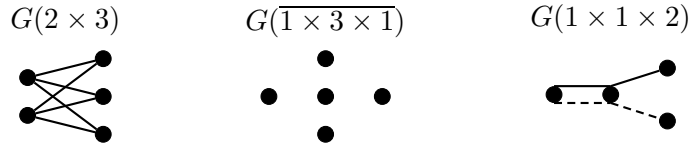
⁶In pictures these lines can be drawn in different styles, to set them apart.

Example 2.3 The WLC game G^* in Example 2.2 has the following game graph:

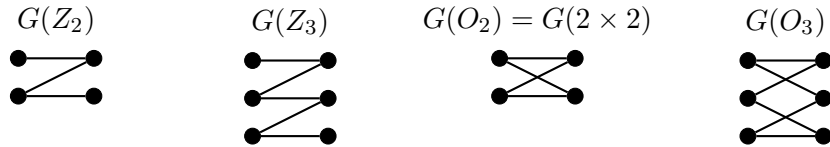


We now define several simple types of WLC games and introduce a uniform notation for them. Since no names of choices are given for these games, each game given here actually corresponds to a class of games with the same structure. However, in this paper we usually consider such games to be equivalent⁷. Let $k_1, \dots, k_n \in \mathbb{N}$.

- $G(k_1 \times \dots \times k_n)$ is the n -player WLC game where the player i has k_i choices and the winning relation is the *universal relation* $C_1 \times \dots \times C_n$.
- $G(\overline{k_1 \times \dots \times k_n})$ is the n -player WLC game where the player i has k_i choices and the winning relation is the *empty relation*. Note that with this notation $G(\overline{k_1 \times \dots \times k_n}) = \overline{G(k_1 \times \dots \times k_n)}$. Some examples:

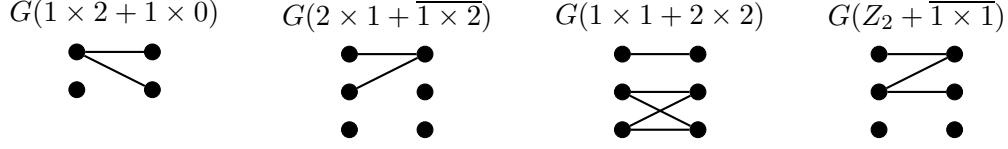


- Let $k \in \mathbb{N}$. We write $G(Z_k)$ for the 2-player WLC game in which both players have k choices and the winning relation forms a single path that goes through all the choices (see below for an example). Similarly, $G(O_k)$, where $k \geq 2$, denotes the 2-player WLC game where the winning relation forms a $2k$ -cycle that goes through all the choices. These are exemplified by the following:

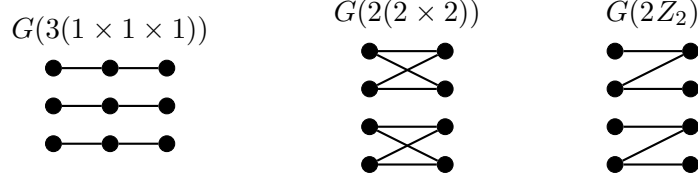


- Suppose that $G(A)$ and $G(B)$ have been defined, both having the same number of players. Then $G(A + B)$ is the *disjoint union* of $G(A)$ and $G(B)$, i.e., the game obtained by assigning to each player a disjoint union of her choices in $G(A)$ and $G(B)$, and where the winning relation for $G(A + B)$ is the union of the winning relations in $G(A)$ and $G(B)$. Some examples:

⁷If a player reasons by pure rationality, the names of the choices should not have an effect on that player's reasoning. We will discuss further this issue later on.



- Let $m \in \mathbb{N}$. Then $G(mA) := G(A + \dots + A)$ (m times). Examples:



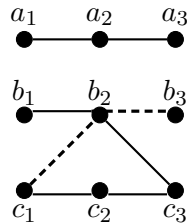
- Recall our “regrouping scenario” (B) from the introduction. If there are n people in the group and there are m possible meeting spots in the city, then the game is of the form $G(m(1^n))$, where $1^n := 1 \times \dots \times 1$ (n times).

2.3 On choice domination and Nash equilibria in WLC games

Given a WLC game $G = (A, C_1, \dots, C_n, W_G)$, we say that a choice $c \in C_i$ is **at least as good as** (respectively, **better than**) a choice $c' \in C_i$ if $W_G^i(c') \subseteq W_G^i(c)$ (respectively, $W_G^i(c') \subsetneq W_G^i(c)$). A choice $c \in C_i$ is **optimal** for a player i if it is at least as good as any other choice of i .

Note that a choice $c \in C_i$ is better than a choice $c' \in C_i$ precisely when c *weakly dominates* c' in the usual game-theoretic sense (see e.g. [17], [23]). Respectively, a choice $c \in C_i$ is an optimal choice of player i when it is a *weakly dominant* choice (i.e., a choice that weakly dominates all other choices).

Note that c *strictly dominates* c' (*ibid.*) if and only if c is a surely winning choice and c' is a surely losing choice. Thus, strict domination between choices is a too strong concept in WLC games. Also, the concept of *Nash equilibrium* (NE) for choice profiles is not very useful here, because not only every winning choice profile is a NE, but so is also every losing choice profile which no player can unilaterally convert to a winning one by changing their choice⁸. For instance, in the game displayed below, the Nash equilibria are not only all 4 winning profiles, (a_1, a_2, a_3) , (b_1, b_2, c_3) , (c_1, b_2, b_3) , and (c_1, c_2, c_3) , but also the losing profiles (a_1, c_2, b_3) and (b_1, c_2, a_3) .



⁸In the special case of two-player games, this latter case amounts to choice profiles consisting only of (surely) losing choices.

We will not make use of Nash equilibria further in this work.

2.4 Symmetries of WLC games and structural protocols

A **protocol** is a mapping Σ that assigns to every pair (G, i) , where G is a WLC game and i a player in G , a nonempty set $\Sigma(G, i) \subseteq C_i$ of choices. Thus, a protocol describes a global nondeterministic strategy for playing any WLC game in the role of any player. Intuitively, such protocol represents a global mode of acting in any situation that involves playing WLC games. Hence, protocols can be informally regarded as global “reasoning styles” or “behaviour modes”. Thus, a protocol can also be identified with an agent who acts according to that protocol in all situations that involve playing different WLC games in different player roles. Note, however, that this correspondence is not bijective, as several agents may behave according to the same protocol and some protocols might not correspond to the behaviour of any (actual) agent.

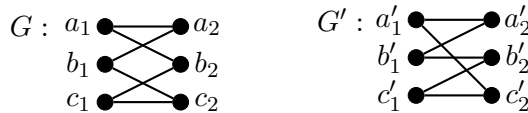
Assuming a setting with no special conventions or preplay communication, a protocol will only take into account the *structural properties* of the game and its winning relation. Thus, it is reasonable to assume that the names of the choices and the names (or ordering) of the players should be of no relevance. In this section we make this issue precise.

Definition 2.4 An isomorphism⁹ between games G and G' is called a **choice-renaming**. An automorphism of G is called a **choice-renaming of G** .

Let $G = (A, C_1, \dots, C_n, W_G)$ be a WLC game. For a player i , we say that the choices $c, c' \in C_i$ are **i -equivalent**, denoted by $c \simeq_i c'$, if there is a choice-renaming of G that maps c to c' . For each $i \leq n$, the relation \simeq_i is an equivalence relation on the set C_i . We denote the equivalence class of $c \in C_i$ by $\llbracket c \rrbracket_i$.

Supposing that a player i does not use names or labels of her choices (or she has no preferences over them), then she should be *indifferent* about the choices that are in the same equivalence class.

Example 2.5 Let $A = \{a_1, b_1, c_1, a_2, b_2, c_2\}$ and $A' = \{a'_1, b'_1, c'_1, a'_2, b'_2, c'_2\}$. Consider the WLC games G and G' whose game graphs are given below.



A function $\pi : A \rightarrow A'$, which maps b_2 to c'_2 , c_2 to b'_2 , and x to x' for all the other choices $x \in A$, is choice-renaming from G to G' . Note that actually both G and G' are of the form $G(O_3)$. A function that maps a_i to b_i , b_i to c_i , and c_i to a_i (for $i \in \{1, 2\}$) is a choice-renaming of G . Therefore $a_1 \simeq_1 b_1 \simeq_1 c_1$ and $a_2 \simeq_2 b_2 \simeq_2 c_2$.

⁹Isomorphism is defined as usual for relational structures (see, e.g., [6]).

Definition 2.6 Consider n -player WLC games

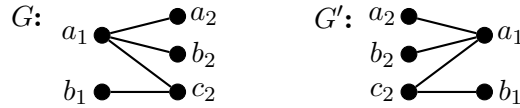
$$G = (A, C_1, \dots, C_n, W_G) \text{ and } G' = (A, C'_1, \dots, C'_n, W'_G).$$

A permutation $\beta : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ is called a **player-renaming** between G and G' if the following conditions hold:

- (1) $C_{\beta(i)} = C'_i$ for each $i \leq n$.
- (2) $W'_G = \{ (c_{\beta(1)}, \dots, c_{\beta(n)}) \mid (c_1, \dots, c_n) \in W_G \}$.

If there is a player-renaming between two WLC games, the games are essentially the same, the only difference being the ordering of the players. Furthermore, the game graph of G' is simply obtained by permuting the columns of the game graph of G .

Example 2.7 Consider the following WLC games:



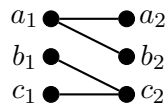
A permutation β , which swaps 1 and 2, is a player-renaming between G and G' .

Definition 2.8 Consider WLC games G and G' . A pair (β, π) is a **full renaming** between G and G' if there is a WLC game G'' such that β is a player-renaming between G and G'' and π is a choice-renaming between G'' and G' . If G and G' have the same domain A , we say that (β, π) is a **full renaming of G** . We say that choices $c \in C_i$ and $c' \in C_j$ in the same game are **structurally equivalent**, denoted by $c \sim c'$, if there is a full renaming (β, π) of G such that $\beta(i) = j$ and $\pi(c) = c'$. It is quite easy to see that \sim is an equivalence relation on the set A of all choices. We denote the equivalence class of a choice c by $[c]$.

We also make the following observations:

- If $c \simeq_i c'$ for some i , then also $c \sim c'$.
- Suppose that there is a sequence G_1, \dots, G_n of WLC games such that for every i there is either a choice-renaming or a player-renaming between G_i and G_{i+1} . Then it is easy to show that there is a full renaming from G_1 to G_n .

Example 2.9 Consider a WLC game of the form $G(1 \times 2 + 2 \times 1)$:



Let β be the permutation which swaps (players) 1 and 2, and let π be the bijection

$$\{(a_1, c_2), (b_1, b_2), (c_1, a_2), (a_2, c_1), (b_2, b_1), (c_2, a_1)\}.$$

Now the pair (β, π) is a full renaming of $G(1 \times 2 + 2 \times 1)$. It is easy to see that \simeq_1 has the equivalence classes $\{a_1\}$ and $\{b_1, c_1\}$, and similarly, \simeq_2 has equivalence classes $\{c_2\}$ and $\{a_2, b_2\}$. Furthermore, \sim has the equivalence classes $\{a_1, c_2\}$ and $\{b_1, c_1, a_2, b_2\}$. Likewise, in the game G^* from Example 2.2 the relation \sim has equivalence classes $\{a_1, b_3\}$, $\{b_1, a_3\}$, $\{a_2, b_2\}$.

We say that a protocol Σ is **structural** if it is “indifferent” with respect to full renamings, which means that, given any WLC games G, G' for which there exists a full renaming (β, π) between G and G' , for any i and any choice $c \in C_i$, it must hold that

$$c \in \Sigma(G, i) \text{ iff } \pi(c) \in \Sigma(G', \beta(i)).$$

Intuitively, this reflects the idea that, when following a structural protocol one acts independently of the names of choices and names (or ordering) of player roles.¹⁰ Thus, following a structural protocol, one cannot tell the difference between choices that are structurally equivalent. Hereafter, unless otherwise specified, we only consider structural protocols.

It is worth noting that if we considered a framework where WLC games were presented so that the names of the choices and players could always be used to uniquely define an ordering (of the players and their choices), solving games could be trivialised by using the pre-negotiated agreement to always choose the lexicographically least tuple from the winning relation. For more on solving coordination games with names or ordering of choices, see [11].

3 Purely rational principles in WLC games

In this section we will analyse various principles which players can follow in WLC games. We will provide justifications for these principles and study which games can be surely won when such principles are followed. Our aim is to characterize which principles are “purely rational” in the sense that all ideally rational agents ought to follow them in all WLC games.

3.1 Purely rational principles

By a **principle** we mean here *any nonempty class of protocols*. Intuitively, these are the protocols “complying” with that principle. Hence principles are simply *properties of protocols* and if protocols are regarded as “reasoning styles”, then principles are properties of

¹⁰In the definition of WLC games, player roles appear as (naturally ordered) indices i . However, this presentation is only for technical convenience, and player roles could instead be called, e.g., “white”, “black” etc.

reasoning. Likewise, if protocols are identified with agents who behave according to them, then principles can be seen as norms which agents follow.

Principles that contain only structural protocols are called **structural principles**. Such principles are properties of structural protocols and therefore they describe behavior which is independent of names of the choices and roles of the players.

We are mainly interested in principles which describe “purely rational and perfect reasoning” that involves neither preplay communication nor conventions. The principles corresponding to such reasoning are defined as follows:

A principle P is called *purely rational* if P is followed by *all ideally rational agents* in *every* WLC game.

An ideally rational agent always reasons in an optimal, faultless way. Note that we *do not give a formal definition* of an ideally rational agent. It is taken to be a philosophical primitive that is central to our study, but yet a notion that we *cannot* define formally in this paper¹¹. Note also that, since purely rational principles are indeed followed by all ideally rational agents, such principles are not based on any particular conventions.

We say that a player i **follows a principle P** in a WLC game G if she plays according to *some* protocol in P . Consider for example the following principles:

$$\begin{aligned} P_1 &:= \{\Sigma \mid \Sigma(G, i) \text{ does not contain any surely losing choices when } W_G \neq \emptyset\}, \\ P_2 &:= \{\Sigma \mid \Sigma(G, i) \text{ contains all choices } c \in C_i \text{ such that } |W_G^i(c)| \\ &\quad \text{is a prime number; if there are no such choices, } \Sigma(G, i) = C_i.\}. \end{aligned}$$

If player i follows P_1 , then she uses some protocol which never selects surely losing choices, if possible. This seems a principle that any rational agent would follow, so it can be regarded as a purely rational principle. Likewise, if player i follows P_2 , then she always plays choices whose degree (in the game graph) is a prime number, if possible. This principle seems arbitrary; it could only be some artificial convention, for example.

We say that a **principle P solves** a WLC game G (or G **is P -solvable**), if G is won whenever every player follows some protocol that belongs to P . Formally, this means that $\Sigma_1(G, 1) \times \cdots \times \Sigma_n(G, n) \subseteq W_G$ for all protocols $\Sigma_1, \dots, \Sigma_n \in P$. The class of all P -solvable games is denoted by $s(P)$.

Hereafter, we will identify (a hierarchy of) principles that can be considered to be purely rational and will analyse the classes of games that they solve. Since we have argued that purely rational principles should be structural, for every principle P which we define we assume that P only consists of *structural* protocols which satisfy the description of P .

¹¹The reader should note that, because the notion of an ideally rational agent is taken as a conceptual primitive, it is problematic to formally prove that any of the principles defined in the paper are actually purely rational. Still, we try to give strong rational justifications for these principles and argue that many of them should be followed by all ideally rational agents. However, when we proceed to stronger principles, it becomes harder to give the principles a solid rational justification and to argue that they would be purely rational. One of our main philosophical aims in the paper is to demonstrate that it is very difficult to identify the boundary of purely rational principles and other principles.

Remark 3.1 When defining principles, we give their definition informally by describing some rational mode of behaviour. For example, the non-losing principle NL in the following section is defined by the sentence “do not play a losing choice, if possible”. We call this description the *intension* of NL, while the *extension* of NL is the actual set of structural protocols that satisfy this intension. In this paper we describe the intensions in natural language but they could be formalized in e.g. some formal logic.

Note that there might be several different intensions that correspond to the same extension. Also, there may be a principle P (consisting of arbitrary protocols) for which there is no natural intension. A similar distinction can also be made for protocols: an agent behaving according to his protocol could have some intension that justifies his behaviour, but we may not be able to identify this intension by only looking at the (extension of the) protocol. This also bears a link to the concept of rationalizable strategies [3], [22].

Remark 3.2 We have defined principles as sets of protocols for mainly conceptual reasons to make a distinction between protocols and their properties. However, for the results of the current paper, we could have defined principles as protocols simply by forming a corresponding “union protocol”. That is, given a principle P, we can form the protocol Σ_P by defining $\Sigma_P(G, i) = \bigcup_{\Sigma \in P} \Sigma(G, i)$, whence we have $s(P) = s(\{\Sigma_P\})$. For the principles defined in this paper, we also happen to have $\Sigma_P \in P$, but generally this does not need to be the case. Moreover, this correspondence is not bijective since we may have $\Sigma_{P_1} = \Sigma_{P_2}$ for some protocols $P_1 \neq P_2$. With the current definitions, we can also combine principles in an easy way by simply taking their intersections—that is, a player follows both P_1 and P_2 if and only if (s)he follows the principle $P_1 \cap P_2$.

3.2 Basic individual rationality

Hereafter we describe principles by the properties of protocols that they determine. We begin by considering the case where players are individually rational, but there is no common knowledge about this being the case. It is safe to assume that any individually rational player would follow at least the following principle.

Fundamental individual rationality (FIR):

Never play a strictly dominated choice¹², if possible.

As noted earlier, strict domination is a very strong concept for WLC games. Following FIR simply means that a player should never prefer a surely losing choice to a surely winning one. Therefore FIR is a very weak principle that can solve only some quite trivial types of games such as $G(1 \times 2 + 1 \times 0)$ (See Figure 1). In general, FIR-solvable games have a simple description: at least one of the players has (at least one) surely winning choice, and all non-winning choices of that player are surely losing. Thus, for example all games of the form $G(k \times l + m \times 0)$ are FIR-solvable.

FIR has two natural strengthenings that can still be considered purely rational:

¹²Recall, that a choice a is strictly dominated by a choice b if the choice b guarantees a strictly higher payoff than the choice a in every play of the game (see e.g. [17], [23]).

1. **Non-losing principle (NL):** Never play a losing choice, if possible.
2. **Sure winning principle (SW):** Always play a winning choice, if possible.

Since losing choices cannot be winning choices, these principles can naturally be put together by taking their *intersection* (recall here that principles are simply sets of protocols that satisfy the given property).

Basic individual rationality (BIR): $NL \cap SW$.

Thus, when following BIR, a player plays a winning choice if she has one, else she plays a non-losing choice. Let us make some observations (see the pictures in Figure 1):

1. NL and SW do not imply each other and neither of them follows from FIR. This can be seen by the following examples.
 - The game $G(1 \times 1 + \overline{1 \times 1})$ is NL-solvable but not SW-solvable. This is because neither of the players has a winning choice, but if each of them chooses their non-losing move they win.
 - The game $G(Z_2)$ is SW-solvable but not NL-solvable. This is because both players have a winning choice, but there are no losing choices. Note that in this game both players can force winning and thus they both would be sure of winning even without knowing that the other player follows SW.
2. FIR-solvable games are solvable by *both* SW and NL. This is because in FIR-solvable games, at least one player i has a winning choice and all the other choices of that player are losing. Hence by following either SW or NL, the player i will select a winning choice.
3. Every BIR-solvable game is *either* NL or SW-solvable. This is because a BIR-solvable game G is won when every player selects a winning choice, if they have one, or else if they each play a non-losing choice. If at least one player has a winning choice in G , then it is SW-solvable, else it is NL-solvable.

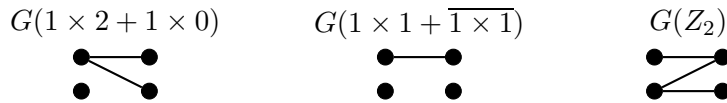
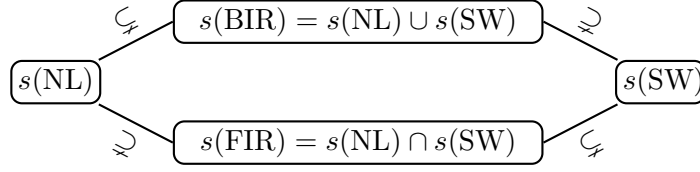


Figure 1: Some BIR-solvable games

Therefore, the sets of games solvable by FIR, NL, SW, BIR form the following lattice:



SW-solvable and NL-solvable games have simple descriptions. In SW-solvable games, at least one player has a surely winning choice. In NL-solvable games, the winning relation forms a nonempty *Cartesian product* between all non-losing choices. BIR-solvable games have (at least) one of these two properties.

Note that in order to follow BIR, the players do not have to make any assumptions on the behavior or rationality of each other. In fact, the players do not even need to know that everyone has a mutual goal in the game; that is, following BIR would be equally rational even in coordination games that are *not cooperative*.

3.3 Common beliefs in rationality and iterated reasoning

In contrast to individual rationality, the collective rationality allows players to make assumptions on each other's rationality. Let P be a (purely rational) principle. When *all players believe that everyone follows P* , they can reason as follows:

- (\star) Suppose that by following P each player i must play a choice from $A_i \subseteq C_i$ (i.e., A_i is the smallest set such that $\Sigma(G, i) \subseteq A_i$ for every $\Sigma \in P$). By this assumption, the players may collectively assume that the actually played game¹³ is $G' := G \upharpoonright (A_1, \dots, A_n)$, and therefore all P -compliant protocols should only prescribe choices in G' .

If players have *common belief* in P being followed, then the reasoning (\star) above can be repeated for the game G' and this iteration can be continued until a fixed point is reached. By $\text{cir}(P)$ we denote the principle of **collective iterated reasoning of P** which prescribes that P is followed in the reduced game obtained by the iterated reasoning of (\star). Note that after every iteration of (\star), the sets of choices for each player become smaller (or remain the same). And since each protocol in any principle P must give nonempty set of choices for any WLC game, $\text{cir}(P)$ cannot make the set of choices empty for any player. From these observations it is easy to see that $s(P) \subseteq s(\text{cir}(P))$ for any principle P .

When considering principles of *collective* rationality further, we will apply collective iterated reasoning of the type described above. It may be argued whether such reasoning counts as purely rational, so a question arises: if P is a purely rational principle, is $\text{cir}(P)$ always purely rational as well? We will not discuss this issue here. We note, however, the extensive literature relating common beliefs and knowledge with individual and collective rationality, see e.g., [7], [16], [21], [29]. See also the following remark on alternative approaches on common belief on a principle P .

¹³In fact, the actually played game may be a proper subgame of G' as the players may also follow other principles.

Remark 3.3 *It is important to note that when applying (\star) , we simultaneously remove all the choices (of all players) that are not admissible by P , and thus the result of the repeated elimination process is unique. However, if we instead eliminated choices of one player at a time, the result of the iterated elimination process could depend on the order in which the players are considered. As we will show in Remark 3.12, certain iteration orders may lead to non-rational results. Hence one could argue that simultaneous elimination of choices is indeed the most rational way of reasoning when having common belief in a principle P .*

Related to collective rationality, consider a situation where player A follows certain principles, say P_1 and P_2 , but (s)he is not sure which principles the other players follow. It is now conceivable that A has a strong reason to believe that all the other players follow P_1 , but A is not sure if these other players also follow P_2 . This could be because P_1 seems rationally obvious while P_2 is only well justified but not followed by all rational players. Alternatively, P_2 could be a more complex principle and A could be skeptical whether the other players are smart enough to follow it. Consider a scenario (for example a specific class of games) where all players reason in the same way so that they have common belief only in P_1 but they in fact all follow both P_1 and P_2 (but no other principles). The games that the players can now solve correspond to games solvable with the principle $\text{cir}(P_1) \cap P_2$.

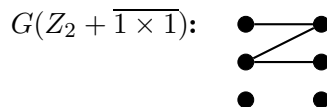
3.4 Basic collective rationality

Here we extend individually rational principles of Section 3.2 by adding common belief in the principles (as described in Section 3.3) to the picture. We first analyse what happens with the principles NL and SW. It is easy to see that the collective iterated reasoning of NL reaches a fixed point in a single step by simply removing the losing choices of every player. Hence $s(\text{NL}) = s(\text{cir}(\text{NL}))$. Collective iterated reasoning with SW also reaches a fixed point in a single step by eliminating all non-winning choices of every player who has a winning choice. But if even one player has a winning choice, then the game is already SW-solvable. Therefore $s(\text{SW}) = s(\text{cir}(\text{SW}))$.

However, even though common belief in NL or SW does not make them stronger by solving more games, there is a difference on the *epistemic level*. For example, the game $G(1 \times 1 + \overline{1 \times 1})$ is solvable with NL even without common belief in NL. But if both players believe that the other player will follow NL, then they will not only win the game, but they also *believe that that game will be won* before it is played.

Assuming common belief in BIR, some games which are not BIR-solvable may become solvable. See the following example.

Example 3.4 *The game $G(Z_2 + \overline{1 \times 1})$ cannot be solved with NL or SW. However, if the players can assume that neither of them selects a losing choice (by NL) and eliminate those choices from the game, then they (both) have a winning choice in the reduced game and can win in it by SW.*



Thus, we define the following principle:

Basic collective rationality (BCR): $\text{cir}(\text{BIR})$.

The example above shows that $s(\text{BIR}) \subsetneq s(\text{BCR})$, i.e., BCR is *stronger* than BIR. The games solvable by BCR have the following characterisation: *after removing all surely losing choices of every player, at least one of the players has a surely winning choice*.

It is worth noting that common belief in SW is not needed for solving games with BCR because a *single* iteration of $\text{cir}(\text{NL})$ suffices. Thus, players could solve BCR-solvable games simply by believing everyone to follow NL, i.e., eliminating all losing moves, and then following SW. By this observation, we have $s(\text{BCR}) = s(\text{cir}(\text{NL}) + \text{SW})$. We also point out that the principle BCR is equivalent to the principle applied in [14] for Strategic Coordination Logic.

3.5 Principles using optimal choices

If a rational player has optimal choices (i.e., at least as good as all other choices), it is natural to assume that she selects such a choice. Note that players may have several optimal choices, or none at all. For example, in the game $G(2 \times 2)$ both players have two optimal choices while in $G(Z_3)$ neither of the players has optimal choices. We now introduce the following principle:

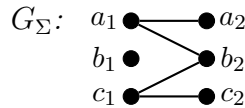
Individual optimal choices (IOC): Play an optimal choice, if possible.

Example 3.5 Recall the WLC game G^* from Example 2.2. For Casper (who is carrying the crowbar) it is a better choice to go to the front door than to the basement. Likewise, for Jonathan (who is carrying the lantern) it is a better choice to go to the basement than to the front door. Therefore the choice a_1 is (the only) optimal choice for player 1 and b_3 is (the only) optimal choice for the player 3. The player 2 (Jesper) does not have any optimal choices, but if both 1 and 3 play their optimal choices, then the game is won regardless of the choice of 2. Therefore, the game G^* is solvable with IOC. But, since no player has winning or losing choices in this game, it is easy to see that it is not BCR-solvable.

Note that if a player has winning choices, then the set of optimal choices is the set of winning choices, and therefore $\text{IOC} \subseteq \text{SW}$. From the description of BIR-solvable games, we see that they are also IOC-solvable.

The next example, together with Example 3.5, shows that IOC is *incomparable* with BCR with respect to the classes of games that are solvable by these principles.

Example 3.6 Consider the following WLC game G_Σ .



By following BCR, player 1 chooses either a_1 or c_1 and player 2 chooses b_2 , whence the game is won. However, G_Σ is not solvable with IOC since player 1 does not have any optimal choices (and may thus end up choosing the losing choice b_1).

As we saw earlier, if a player does not have optimal choices, following only IOC might lead to playing a losing choice. In order to avoid pathological cases like this, we should at least add NL to IOC.

Improved basic individual rationality (BIR⁺): $\text{IOC} \cap \text{NL}$

Since $\text{IOC} \subseteq \text{SW}$, we have $\text{BIR}^+ \subseteq \text{BIR}$. Note that, unlike BCR, the principle BIR⁺ is only based on *individual* reasoning. However, BIR⁺ is nevertheless stronger than BCR as shown by the following proposition.

Proposition 3.7 $s(\text{BCR}) \subsetneq s(\text{BIR}^+)$.

Proof Suppose first that $G \in s(\text{BCR})$. Then each player i has a nonempty set A_i of winning choices in the reduced game after removing all losing choices of all the other players. But now every choice in A_i must be an optimal choice of i in the original game G . Hence, by following BIR⁺, the player i will play a choice from A_i (by IOC) while all the other players play a non-losing choice (by NL), whence the game is won. Therefore $G \in s(\text{BIR}^+)$ and thus $s(\text{BCR}) \subseteq s(\text{BIR}^+)$. In Example 3.5 we saw that G^* is solvable with IOC but not with BCR. Therefore $s(\text{BCR}) \subsetneq s(\text{BIR}^+)$. \square

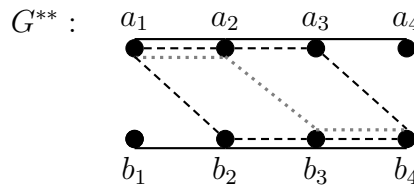
We now consider the collective version of IOC:

Collective optimal choices (COC): $\text{cir}(\text{IOC})$

Proposition 3.8 $s(\text{BIR}^+) \subsetneq s(\text{COC})$.

Proof We first show that $s(\text{BIR}^+) \subseteq s(\text{COC})$. Suppose that a WLC game G is BIR⁺-solvable, i.e., the game is won when every player plays an optimal choice, if they have any, else they play a non-losing choice. Let G' be the game that is obtained after the first collective iteration of $\text{cir}(\text{IOC})$. Now, all the remaining non-losing choices of every player in G' must be winning choices. Since winning choices are also optimal choices, all losing choices are eliminated in the second iteration of $\text{cir}(\text{IOC})$. After that, all combinations of the remaining choices are winning. Thus, the game is won by following COC.

Now, consider the following WLC game G^{**} .



Here only players 1 and 4 have optimal choices a_1 and b_4 , respectively, and no player has losing choices. Hence we see that G cannot be solved with BIR^+ . (By following BIR^+ , players may end up selecting the choice profile (a_1, b_2, a_3, b_4) which is not winning.) However, after the first iteration of $\text{cir}(\text{IOC})$, the players 2 and 3 have optimal choices a_2 and b_3 , respectively. Hence, by following COC , the players end up choosing the winning choice profile (a_1, a_2, b_3, b_4) . (Note that we can construct a similar game for $2n$ players, where it takes n iterations of $\text{cir}(\text{IOC})$ for solving it.) \square

Now, let us consider what happens in the special case of 2-player WLC-games. We first observe that *the only optimal choices in a 2-player WLC game G (where $W_G \neq \emptyset$) are those that are winning against all non-surely losing choices of the other player*. Consequently, when considering 2-player WLC games, we have $s(\text{IOC}) \cup s(\text{COC}) \subseteq s(\text{BCR})$. By combining this with the results of Propositions 3.7 and 3.8, we obtain the following result.

Proposition 3.9 *For 2-player WLC games: $s(\text{BCR}) = s(\text{BIR}^+) = s(\text{COC})$.*

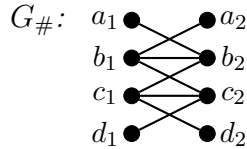
3.6 Elimination of weakly dominated choices

In game-theory, rationality is usually associated with elimination of dominated strategies. As noted in Section 3.2, *strict* domination is a too strong concept for WLC-games. Weak domination, on the other hand, gives the following principle when applied individually.

Individually rational choices (IRC): Do not play a choice a when there is a better choice b available. That is, if $W_G^i(a) \subsetneq W_G^i(b)$, then the player i should not play a .

Note that by the definition, $\text{IRC} \subseteq \text{NL} \cap \text{IOC}$ and therefore $s(\text{BIR}^+) \subseteq s(\text{IRC})$. The inclusion here is proper since there are WLC games that are solvable with IRC but not with BIR^+ .

Example 3.10 *Consider the following WLC game $G_\#$.*



In $G_\#$, all players have neither losing choices nor optimal choices, and therefore it cannot be solved with BIR^+ . But the choice b_1 is better than a_1 , and likewise c_1 is better than d_1 . (Note that b_1 and c_1 are not comparable with each other.) Therefore, by following IRC, the player 1 will play either b_1 or c_1 . With the same reasoning, the player 2 will play either b_2 or c_2 , which leads to a win. Therefore $G_\# \in s(\text{IRC})$.

The COC -solvable game G^{**} (in the proof of Proposition 3.8) is not solvable with IRC. (This is because neither of the moves a_2 and b_2 (respectively a_3 and b_3) is better than the other.) On the other hand, the game $G_\#$ in Example 3.10 is not solvable with COC , and

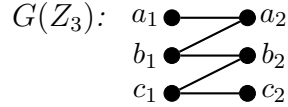
therefore IRC is incomparable with COC in the general case. However, in the 2-player case $s(\text{COC}) \subsetneq s(\text{IRC})$, since then $s(\text{COC}) = s(\text{BIR}^+)$ by Proposition 3.9.

We next assume common belief in IRC. As commonly known (see e.g. [21]), *iterated elimination of weakly dominated strategies* eventually stabilises in some reduced game but different elimination orders may produce different results (cf. Remark 3.12). However, when applying $\text{cir}(\text{IRC})$, the process will stabilise to a unique reduced game since all weakly dominated choices are always removed *simultaneously*. By following the next principle, players will play a choice within this reduced game.

Collective rational choices (CRC): $\text{cir}(\text{IRC})$

The following example shows that $s(\text{IRC}) \subsetneq s(\text{CRC})$.

Example 3.11 Consider the following WLC game.



We first note that b_1 is a better choice than a_1 and likewise b_2 is a better choice than c_2 . Therefore, by following IRC, player 1 will play a choice from $\{b_1, c_1\}$ and player 2 will play from $\{a_2, b_2\}$, which does not guarantee winning. However, after eliminating a_1 and c_2 , then b_1 is better than c_1 and b_2 is better than a_2 . Thus, by following CRC and doing one more iteration of $\text{cir}(\text{IRC})$, player 1 and 2 have only the choices b_1 and b_2 which are winning.

In $G(Z_3)$, we needed two iterations of $\text{cir}(\text{IRC})$. It is easy to see that in the game $G(Z_4)$ the iterations are done analogously and the fixed point is reached in 3 iterations. Likewise, we can see that $n - 1$ iterations of $\text{cir}(\text{IRC})$ are needed for solving $G(Z_n)$. Therefore, the numbers of iterations of $\text{cir}(\text{IRC})$ form a proper hierarchy of CRC-solvable 2-player WLC games.

Remark 3.12 As discussed in Remark 3.3, if the iterated process of choice elimination is performed in a non-simultaneous fashion, the resulting reduced game might be different depending on the selected elimination order. This can easily lead to reasoning patterns that may appear sound, but are arguably irrational. For example, in the game $G(Z_3)$ above, we could first eliminate the weakly dominated choice a_1 of player 1 and then the choices a_2 and c_2 of player 2, thereby ending up with the subgame with the choices b_1, b_2, c_1 remaining. The player 1 could then conclude that b_1 and c_1 are equally good choices for him. (Note that neither of these choices is weakly dominated in the original game.) Symmetrically, we could first eliminate the choice c_2 of player 2 and then the choices a_1 and c_1 of player 1, ending up with the game with the choices b_1, a_2, b_2 . Now player 2 could conclude that the choices a_2 and b_2 are equally good for him. This way player 1 could end up choosing c_1 (based on the first reduced game) and player 2 choice a_2 (based on the second reduced game), whence the players would not coordinate.

For an example of even more problematic reasoning leading to irrational results, consider the SW-solvable game $G(Z_2)$. Here player 1 could first assume that player 2 is following SW and eliminate the non-surely winning choice of player 2. Since the remaining choices of player 1 are surely winning in the reduced game, (s)he could then choose either of them. But similarly, player 2 could assume that player 1 is not playing the non-surely winning choice whence it would be fine (for player 2) to play any choice. As a result, the players could end up choosing a non-winning choice profile and lose $G(Z_2)$. Since the reasoning here is clearly faulty, this demonstrates that the elimination of choices “blindly” in an arbitrary order can lead to irrational behavior.

3.7 Symmetry-based principles

By *only* following the concept of rationality from game-theory, one *could* argue that CRC reaches the border of rational principles. However, we now will define more principles which are incomparable with CRC but can still be regarded as purely rational, and hence to be followed by all ideally rational players. These principles are based on *symmetries* in WLC games and the assumption that players follow only structural protocols is central here.

We begin with auxiliary definitions. We say that a choice profile (c_1, \dots, c_n) **exhibits a bad choice symmetry** if $\llbracket c_1 \rrbracket_1 \times \dots \times \llbracket c_n \rrbracket_n \not\subseteq W_G$ (recall Definition 2.4), and that a choice c **generates a bad choice symmetry** if σ_c exhibits bad choice symmetry for *every* choice profile σ_c that contains c .

Elimination of bad choice symmetries (ECS):

Never play choices that generate a bad choice symmetry, if possible.

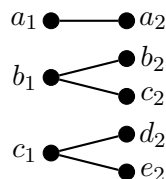
Why should this principle be considered rational? Suppose that a player i plays a choice c_i which generates a bad choice symmetry. It is now possible to win only if some tuple

$$(c_1, \dots, c_{i-1}, c_i, c_{i+1}, \dots, c_n) \in W_G$$

is eventually chosen. However, due to structural symmetry, the players have *exactly the same reasons* to play in such a way that any other tuple in $\llbracket c_1 \rrbracket_1 \times \dots \times \llbracket c_n \rrbracket_n$ is selected—and that other tuple may possibly be a losing one since $\llbracket c_1 \rrbracket_1 \times \dots \times \llbracket c_n \rrbracket_n \not\subseteq W_G$.

For a typical example of using ECS, suppose that the game graph of G has two (or more) connected components that are isomorphic to each other. Since no player can detect a difference between those components, all players should avoid playing choices from them.

Example 3.13 Consider the WLC game $G(1 \times 1 + 2(1 \times 2))$:



In this game $b_1 \simeq_1 c_1$ and $b_2 \simeq_2 c_2 \simeq_2 d_2 \simeq_2 e_2$. Since all the choice profiles in $\{b_1, c_1\} \times \{b_2, c_2, d_2, e_2\}$ are not winning, we see that both b_1 and c_1 generate a bad choice symmetry. Likewise, b_2, c_2, d_2 and e_2 generate a bad choice symmetry. Therefore, by following ECS, the players will choose a_1 and a_2 .

While ECS only considers symmetries between similar choices, the next principle takes symmetries *between players* into account. Consider a choice profile $\vec{c} = (c_1, \dots, c_n)$ and let $S_i^p(\vec{c}) := \{c_i\} \cup (C_i \cap \bigcup_{j \neq i} [c_j])$ for each i (recall Definition 2.8). We say that (c_1, \dots, c_n) **exhibits a bad player symmetry** if $S_1^p(\vec{c}) \times \dots \times S_n^p(\vec{c}) \not\subseteq W_G$ and a choice c **generates a bad player symmetry** if σ_c exhibits a bad player symmetry for every choice profile σ_c that contains c .

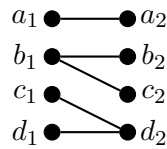
Elimination of bad player symmetries (EPS):

Never play choices that generate bad player symmetries, if possible.

Here the players assume that all players reason similarly, or alternatively, each player wants to play so that she would at least coordinate with herself in the case she was to use her protocol to make a choice in each player role of a WLC game. Suppose that the players have some reasons to select a choice profile (c_1, \dots, c_n) . Now, if there are players $i \neq j$ and a choice $c'_j \in C_j$ such that $c'_j \sim c_i$, then the player j should have the same reason to play c'_j as i has for playing c_i . Hence, if the players have their reasons to play (c_1, \dots, c_n) , they should have the same reasons to play any choice profile in $S_1^p(\vec{c}) \times \dots \times S_n^p(\vec{c})$. Winning is not guaranteed if $S_1^p(\vec{c}) \times \dots \times S_n^p(\vec{c}) \not\subseteq W_G$.

It is worth noting that EPS bears a close resemblance to the notion of *superrationality* defined by Hofstadter [24].

Example 3.14 Consider EPS in the case of a two-player game WLC game G . If for a given choice $c \in C_1$, there is a structurally equivalent choice $c' \in C_2$ such that $(c, c') \notin W_G$, then by following EPS, player 1 does not play the choice c (and likewise player 2 does not play the choice c'). With this kind of reasoning, some CRC-unsolvable games like $G(1 \times 1 + 1 \times 2 + 2 \times 1)$ below become solvable.



Note also that the game G^* from Example 2.2 is EPS-solvable since both choices b_1 and a_3 generate a bad player symmetry.

Example 3.15 In Example 3.11 we showed that in order to solve $G(Z_n)$ by CRC it takes $n - 1$ collective iterations and after that the “middle choices” are selected by both of the players. The game $G(Z_n)$ can also be solved by EPS and the players will end up choosing the same choices as with CRC. This is because every other choice—except the middle choice—generates a bad player symmetry.

Finally, we introduce a principle that takes both types of symmetries into account. For a choice profile $\vec{c} = (c_1, \dots, c_n)$ let $S_i(\vec{c}) := C_i \cap \bigcup_j [c_j]$ for each i . We say that (c_1, \dots, c_n) **exhibits a bad symmetry** if $S_1(\vec{c}) \times \dots \times S_n(\vec{c}) \not\subseteq W_G$, and a choice c **generates a bad symmetry** if σ_c exhibits a bad symmetry for every choice profile σ_c that contains c .

Elimination of bad symmetries (ES):

Never play choices that generate bad symmetries, if possible.

It is easy to see from the definition of bad symmetry that if a choice c generates either a bad choice symmetry or a bad player symmetry, then c also generates a bad symmetry. Therefore, by using Claim I from Section 5, it is easy to show that $s(\text{ECS}) \cup s(\text{EPS}) \subseteq s(\text{ES})$.

By the definitions of ECS and EPS, it is clear that they can solve all NL-solvable games and therefore also ES can solve all NL-solvable games. Furthermore, we can show that the classes of games solvable by ECS, EPS, and CRC are *completely independent* of each other. See the table in Figure 2.



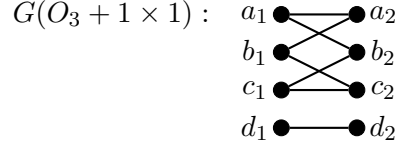
Class of games \mathcal{G}	Example of a game in the class \mathcal{G}
$s(\text{ECS}) \setminus (s(\text{EPS}) \cup s(\text{CRC}))$	$G(1 \times 1 + 2(1 \times 2))$
$s(\text{EPS}) \setminus (s(\text{ECS}) \cup s(\text{CRC}))$	$G(1 \times 1 + 1 \times 2 + 2 \times 1)$
$s(\text{CRC}) \setminus (s(\text{ECS}) \cup s(\text{EPS}))$	
$(s(\text{ECS}) \cap s(\text{EPS})) \setminus s(\text{CRC})$	$G(1 \times 1 + 2(2 \times 2))$
$(s(\text{ECS}) \cap s(\text{CRC})) \setminus s(\text{EPS})$	
$(s(\text{EPS}) \cap s(\text{CRC})) \setminus s(\text{ECS})$	$G(Z_3)$

Figure 2: Mutual independence of the principles ECS, EPS, and CRC.

The WLC game given on Figure 2 in the class $s(\text{CRC}) \setminus (s(\text{ECS}) \cup s(\text{EPS}))$ is also unsolvable with ES and therefore ES and CRC are incomparable with each other. This particular game is also SW-solvable, and thus it follows that all symmetry based principles are incomparable with SW. Since ECS and EPS are incomparable and $s(\text{ECS}), s(\text{EPS}) \subseteq s(\text{ES})$, it also follows that ES is stronger than both ECS and EPS.

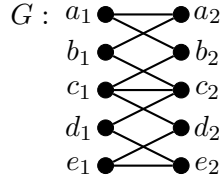
So far we have only presented examples of such ECS-solvable games that contain isomorphic connected components. In the following example, we see how ECS can be used for eliminating moves from a single component. This particular example can also be solved with EPS (and ES) but not with CRC.

Example 3.16 *In the WLC game $G(O_3 + 1 \times 1)$, there are no weakly dominated choices and thus it is not CRC-solvable. However, by applying ECS, EPS or ES, players will play choices d_1 and d_2 which are winning.*



Symmetry principles can solve many games whose game graphs consist of several connected components. It is easy to see that none of these kinds of games can be solved with CRC nor with any other principles we have presented in this paper. However, when considering games whose graph is connected, it is not obvious whether symmetry principles could solve any more games than CRC. In the next example we show that, indeed, there are games, whose graph consists only of a single component, such that symmetry principles solve them, but they are not solvable by CRC (nor any other principles presented so far).

Example 3.17 *In the WLC game G below, there are no weakly dominated choices. However, by applying ECS, EPS or ES, players will pick choices c_1 and c_2 which are winning. (Note here that G is almost of the type of $G(O_5)$, the only difference being a single extra edge that “forms a diagonal of the 10-cycle”.)*



3.8 Compatibility of ES and CRC

Recall that ES is the strongest of the symmetry principles and CRC is the strongest of all the other principles presented in this paper. Since they are incomparable (with respect to solvable games), it is natural to ask whether they can be combined. We first prove the following lemma about the relationship between IRC and ES.

Lemma 3.18 *Let $G = (A, C_1, \dots, C_n, W_G)$ be WLC game and let $c \in C_i$ for some $i \leq n$ be a choice that does not generate a bad symmetry. Then none of the choices $d \in C_i$ which are better than c can generate a bad symmetry.*

Proof For the sake of contradiction, assume that there is a choice $d \in C_i$ which generates bad symmetry and which is better than c . We first observe that the following holds (by the definition of structural equivalence).

$$\begin{aligned} \text{If } d \sim d' \text{ for some } d' \in C_j \ (j \leq n), \text{ then there is a choice } c' \in C_j \quad (\star) \\ \text{such that } c \sim c' \text{ and } d' \text{ is better } c'. \end{aligned}$$

Because c does not generate a bad symmetry, there is a choice profile \vec{c} of the form $(c_1, \dots, c, \dots, c_n)$ which does not exhibit a bad symmetry, i.e. $S_1(\vec{c}) \times \dots \times S_n(\vec{c}) \subseteq W_G$. In

particular $\vec{c} \in W_G$. Let $\vec{d} := (c_1, \dots, d, \dots, c_n)$. Since d is better than c , we have $\vec{d} \in W_G$. Because d generates a bad symmetry, \vec{d} must exhibit a bad symmetry. Therefore there is a choice profile $\vec{e} \in S_1(\vec{d}) \times \dots \times S_n(\vec{d})$ such that $\vec{e} \notin W_G$.

Now \vec{e} must contain at least one choice d' , for which $d \sim d'$, as otherwise we would have $\vec{e} \in S_1(\vec{c}) \times \dots \times S_n(\vec{c}) \subseteq W_G$. Now, by (\star) , for every such choice d' , there is a choice c' such that $c \sim c'$ and d' is better than c' . Let $\vec{e}[c'/d']$ be a choice profile which is obtained from \vec{e} by replacing every d' , for which $d \sim d'$, with a corresponding choice c' . Since $\vec{e} \notin W_G$ and all choices d' are better than the corresponding choice c' , we must have $\vec{e}[c'/d'] \notin W_G$. But this is a contradiction since $\vec{e}[c'/d'] \in S_1(\vec{c}) \times \dots \times S_n(\vec{c}) \subseteq W_G$. \square

We are now ready to show that ES and CRC can be combined simply by taking their intersection.

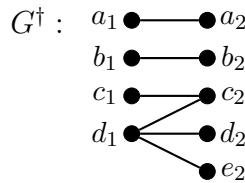
Proposition 3.19 *In every WLC game it is possible to follow both the principle ES and the principle CRC, i.e., $\text{CRC} \cap \text{ES} \neq \emptyset$.*

Proof Let G be a WLC game and let i be a player in G . We first note that if all choices of i generate a bad symmetry, then the principle ES does not set any restrictions and thus, by following CRC, the player i also trivially follows $\text{CRC} \cap \text{ES}$.

Suppose then that i has at least one choice c which does not generate a bad symmetry. Now, by Lemma 3.18, there cannot be any choice d which is better than c such that d generates a bad symmetry. Therefore it follows, by induction on the steps of the choice elimination process, that the iteration of IRC cannot remove all choices of i which do not generate bad symmetry. Hence in the reduced game, obtained by CRC, there must still be at least one choice which does not generate a bad symmetry. Now i will follow $\text{CRC} \cap \text{ES}$ by selecting any such choice. \square

As shown by the following example, $s(\text{CRC}) \cup s(\text{ES}) \subsetneq s(\text{CRC} \cap \text{ES})$. And therefore $\text{CRC} \cap \text{ES}$ is the strongest of all the principles we have presented so far.

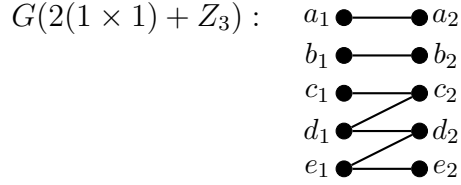
Example 3.20 *Consider the following WLC game:*



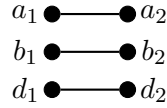
By following CRC in G^\dagger , the player 1 will not play the choice c_1 and the player 2 will not play d_2 or e_2 . And by following ES, the player 1 will play neither the choice a_1 nor b_1 and likewise the player 2 will play neither the choice a_2 nor b_2 . Therefore G^\dagger cannot be surely won by following either CRC or ES alone. But by following $\text{CRC} \cap \text{ES}$, the player 1 will select d_1 and the player 2 will select c_2 , whence the game is surely won.

It is important to note that, by following $\text{CRC} \cap \text{ES}$, the principle ES is applied to the *original* game and not to the reduced game that is obtained by iterating IRC. If we applied ES to the reduced game instead, this would lead to a different kind of principle that allows (arguably) irrational choices. See the following example.

Example 3.21 *Consider the following WLC game:*



By iterating IRC on $G(2(1 \times 1) + Z_3)$, the choices c_1 , c_2 , e_1 and e_2 are eliminated and we obtain the following reduced game:



Since all the choices in this reduced game generate a bad symmetry, the game is not solvable by ES. However, by following $\text{CRC} \cap \text{ES}$ in $G(2(1 \times 1) + Z_3)$, the players will select the choice profile (d_1, d_2) and win. (In this particular example, the same result is obtained by following ES alone as all the other choice profiles exhibit a bad symmetry.)

Notice that the choice profile (d_1, d_2) does not exhibit a bad symmetry in the original game, but it exhibits a bad symmetry in the reduced game after the iteration of IRC. By this example, we claim that it is questionable to apply ES in any reduced game that is obtained by eliminating choices that are excluded by other principles.

4 On the limits of pure rationality

How far can we go up the hierarchy of purely rational principles? This seems a genuinely difficult question. In this section we will first study the compatibility of purely rational principles and then present two simple principles which are naturally justified, but go beyond pure rationality. Finally we will present a complete hierarchy (with respect to solvable games) of all principles defined in this paper and discuss whether one can identify a strongest purely rational principle.

4.1 Merging purely rational principles

Since we have defined that purely rational principles are followed by all ideally rational players, it is clear that purely rational principles are closed under intersections. That is, if P_1 and P_2 are purely rational, then $P_1 \cap P_2$ is purely rational, too. This gives us a method

to identify which principles cannot be purely rational: if P_1 or P_2 are **incompatible**, in sense that $P_1 \cap P_2 = \emptyset$, then it must be that at least one of P_1 and P_2 is not purely rational. However, even if the (extensions of) P_1 and P_2 are incompatible in this way, sometimes it may be that the intensions (recall Remark 3.1) behind both P_1 and P_2 seem rational. See the following example.

Example 4.1 *For each positive integer k , consider the following principle P_k :*

If the game contains two or more copies of $G(k \times k)$ as disjoint components, then do not choose from these components, if possible.

This may seem a purely rational principle for all k , as it is obvious that if there are two or more disjoint components $k \times k$ in a game, then choosing from such a component means that a win cannot be guaranteed due to the same symmetry-based idea as the one used in the justification of ECS. However, for example P_2 and P_3 are not compatible, as for example in the game $G(2(2 \times 2) + 2(3 \times 3))$, following both P_2 and P_3 would leave the players with no choices at all. (Note, however, that the symmetry principle ECS deals with all P_k , for all k , simultaneously).

Even if the extensions of the principles P_1 and P_2 are incompatible, it is still possible that the informal background *intuition* of each of their *intensions* is rational and thus P_1 and P_2 can be combined in some natural way by taking both of the intensions into account. Suppose, e.g., that the intension of P_1 is “do not play choices of type A, if possible” and the intension of P_2 is “do not play choices of type B, if possible” (as in the example above). Now we can naturally combine these intensions into a new intension: “do not play choices of type A or B, if possible”. The resulting principle will now prescribe *any choice* in those games where all choices are “bad” due to being either of type A or B.

From these observations we point out that we may not be sure that (the extension of) a principle P is purely rational unless we know that it is compatible with all purely rational principles. Nevertheless, we may still know that the intuition behind the intension of P is rationally justified.

4.2 Probabilistic reasoning vs. Occam razor

We now mention—without providing precise formal definitions—two structural principles for which it may seem somewhat controversial to claim them purely rational in our sense, but they are definitely meaningful and natural nevertheless.

The first one is the **principle of probabilistically optimal reasoning (PR)**. Informally put, this principle prescribes to always play a choice that has as large winning extension as possible. Such choices have the highest probability of winning, supposing that all the other players play randomly (but *not* if the others follow PR, too: consider e.g. $G(1 \times 2 + 2 \times 1)$). Note, however, that following PR can violate the symmetry principles, as demonstrated by the game $G(2(2 \times 2) + 1 \times 1)$. So, its application should be suitably restricted, as part of its formulation, to be only applied when none of the so far discussed principles of pure rationality apply.

With PR one can solve games like $G(1 \times 1 + 2 \times 2)$ that are unsolvable with all other principles presented here. However, in $G(1 \times 1 + 2 \times 2)$ one could also reason (perhaps less convincingly) that both players should pick their choices from the subgame $G(1 \times 1)$ since that is the ‘*simplest*’ (and, also the only ‘*unique*’) winning choice profile. We call this kind of reasoning the **Occam razor principle (OR)**. Its intuitive and inevitably imprecise formulation is “If there is one simplest way to coordinate, make that choice.” This principle relates to the idea of *focal points* [20], [25], [28]. Different technically precise formulations are possible, but it is not easy to see which one of them (if any) would best capture the spirit of the principle.

Note that $G(1 \times 1 + 2 \times 2)$ can be won if both players follow PR or if both follow OR, but not if one follows PR while the other follows OR. Moreover, in this game it is impossible for a player to follow *both* PR and OR. Hence, *at least one* of these principles is not purely rational. Actually, it can be argued that *neither of them* is purely rational.

Remark 4.2 *Consider a setting where players have good reasons to assume that all the other players are playing randomly. (One could argue that they can make this assumption “by default” if there is no common belief in rationality.) Now, if a player does not have a winning choice, then winning is not guaranteed. Therefore it makes sense to “optimize the expectation of winning” and therefore follow the principle PR. If all the players happen to reason this way, then games like $G(1 \times 1 + 2 \times 2)$ are in fact won.*

However, if we assume common belief in rationality, then it seems obvious that the players can no longer assume that the other players play randomly. Therefore the justification for following PR becomes questionable, and thus, it is no more clear whether following PR is rational in games like $G(1 \times 1 + 2 \times 2)$. One could even argue, based on this observation, that common belief in rationality can sometimes be harmful for the players. This is because the use of PR was based on the assumption that the other players are playing randomly, and this is no longer the case if they are assumed to be rational.

4.3 Hierarchy of rationality principles with respect to solvable games

The partially ordered diagram in Figure 3 presents the hierarchy of solvable games with the rationality principles that we have presented in this paper. The principles that only use individual reasoning are put in normal (single) frames and the ones that use collective reasoning have double frames. Note that the diagram in Figure 3 is complete in the sense that no new lines can be added to it (neither in the general nor in the 2-player case).

It is natural to ask whether there exists a *strongest* purely rational principle. That is, a principle P which is followed by all ideally rational players and which can solve more games than any other purely rational principle. Such a principle indeed exists, since we can simply take the *intersection* of *all* purely rational principles to obtain a principle which is purely rational and which can solve all games that are solvable by some purely rational principle. Alternatively, we obtain the (same) strongest purely rational principle by forming a set of all *protocols* corresponding to all ideally rational players (recall that rational players can be identified with the protocols that they follow).

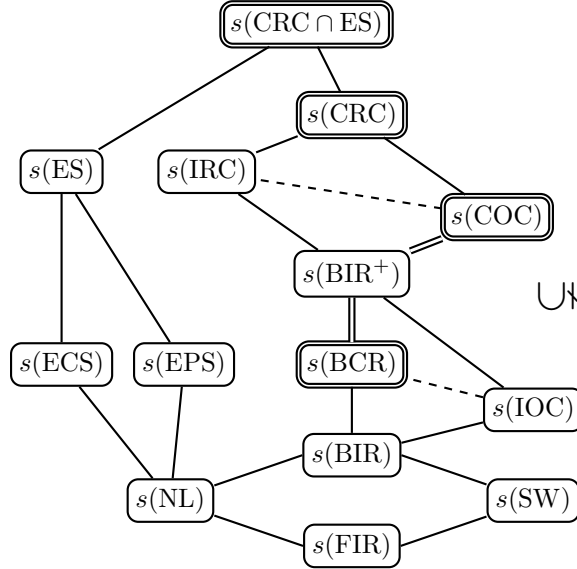


Figure 3: Hierarchy of rationality principles with respect to solvable games. Normal lines represent proper inclusions in both the general *and* 2-player case. *Double* lines represent proper inclusions in the general case—in the 2-player case there is an identity. *Dashed* lines represent proper inclusions in the 2-player case—in the general case the two sets are not comparable.

However, even though there really exists a strongest purely rational principle, we cannot define it explicitly, unless we know all the ideally rational players (which we have taken as a primitive notion). Of the principles presented in this paper, $\text{CRC} \cap \text{ES}$ is the strongest one which could be claimed to be purely rational, but we leave it open whether this principle could be strengthened any further by intersecting it with some other purely rational principles¹⁴.

5 Coordination with structural conventions and structurally unsolvable games

So far we have presented several principles with different levels of justification for being purely rational and therefore naturally applicable without any preplay communication between the players. However, in general, many structural principles lack a rational justification and they may even look completely arbitrary. Such principles can be called *structural conventions*. Even though possibly arbitrary, structural conventions could nevertheless be explicitly agreed upon by a group of players. Alternatively, they could also be considered to

¹⁴It is also conceivable that $\text{CRC} \cap \text{ES}$ may turn out not purely rational due to being incompatible with some other principle that is found to be purely rational. But even if this was the case, we could still argue that the *informal background intuitions* behind CRC and ES are purely rational, even if their formal extensions (i.e., the corresponding sets of protocols) are not (cf. Section 4.1).

emerge naturally with little or no explicit negotiations. (Even highly arbitrary conventions could conceivably be followed in different, possibly hypothetical, communities or cultures.)

We now discuss briefly how WLC games can be solved if the players can negotiate (communicate) *before* they are presented with the WLC game they are to play. We make the *assumption* that via preplay communication, the players can agree on which *structural principles* they will use in the game. Thus the assumption here is that the players can discuss the details of different games only up to structural equivalence. This assumption is natural, e.g., in scenarios where there is no obvious reason for preferring some names of potential game choices with the expense of others. It can even be plainly impossible to distinguish structurally equivalent choices. In these kinds of scenarios it is natural to assume reasoning only up to structural equivalence.

Even when entirely arbitrary structural principles (i.e., structural conventions) are allowed, there exist WLC games that cannot be solved. Games that cannot be solved with *any* structural principle are called **structurally unsolvable**. Probably the simplest non-trivial example of such a game is $G(2(1 \times 1))$. Next we will characterise the class of all structurally unsolvable WLC games.

We say that G is **structurally indeterminate** if all choice profiles in W_G exhibit a bad symmetry (recall the definition of the principle ES). For example, the game $G(1 \times 2 + 2 \times 1 + 3 \times 3 + 3 \times 3)$ is structurally indeterminate because every choice profile exhibits a bad symmetry. On the other hand, the game $G(1 \times 2 + 2 \times 1 + 3 \times 3 + 4 \times 4)$ is *not* structurally indeterminate, as none of the winning choice profiles with nodes of degree 3 or 4 exhibit a bad symmetry.

Theorem 5.1 *No structural principle can solve a structurally indeterminate game.*

Proof For the sake of contradiction, suppose that there is a structural principle P and a structurally indeterminate WLC game G such that $G \in s(P)$. Let Σ be any protocol in the principle P . Since P is a structural principle, Σ must be a structural protocol. Since $P' \subseteq P$ implies $s(P) \subseteq s(P')$, the also the singleton principle $\{\Sigma\}$ solves G .

Let $(u_1, \dots, u_n) \in \Sigma(G, 1) \times \dots \times \Sigma(G, n)$. Since G is structurally indeterminate, (u_1, \dots, u_n) must exhibit a global losing symmetry. Therefore there is a choice profile $(u'_1, \dots, u'_n) \in U_1 \times \dots \times U_n$ such that $(u'_1, \dots, u'_n) \notin W_G$. Since Σ is a structural protocol, we must have $(u'_1, \dots, u'_n) \in \Sigma(G, 1) \times \dots \times \Sigma(G, n)$. Since $(u'_1, \dots, u'_n) \notin W_G$, we have $\Sigma(G, 1) \times \dots \times \Sigma(G, n) \not\subseteq W_G$. Therefore $\{\Sigma\}$ does not solve G , which is a contradiction. \square

The characterisation given by the theorem above is optimal in the sense that all games that are not structurally indeterminate, can be solved by some structural principle. This follows from the following even stronger result.

Theorem 5.2 *There exists a protocol Σ such that the principle $\{\Sigma\}$ solves all WLC games that are not structurally indeterminate.*

Proof The idea is simply to define a protocol that chooses, on an arbitrary input $(G, i) = ((A, C_1, \dots, C_n, W_G), i)$, where G is not structurally indeterminate, a node from a tuple of W_G that does not exhibit global losing symmetry. The only difficulty is that there may be several such tuples in G , and these tuples do not necessarily form a Cartesian product. We next briefly describe how to circumvent this problem. Informally, we will just consider different linear orderings of the input structure (as well as structures obtained from it by a player renaming) and choose the lexicographically smallest suitable tuple.

Firstly, we use the encoding of relational structures by binary strings given in Chapter 6 of [18]. Within this standard encoding scheme, every string encoding of a structure requires a linear ordering of the domain of the encoded structure. The linear order then defines a lexicographic order of the tuples of the structure. The order is used for encoding each relation C_i (with C_1 first and C_n last) and also W_G . We note that a single structure will typically have several encodings, as different linear orderings of the domain can define different encoding strings. Non-isomorphic structures will never share the same encoding.

Now, given an input $(G, i) = ((A, C_1, \dots, C_n, W_G), i)$, we do the following. We first define the finite set \mathcal{G} that contains exactly all structures that can be obtained from G by a player renaming, including G itself. We will below refer to the structures in \mathcal{G} as *renamings of G* . We note that no two structures in \mathcal{G} are isomorphic, but all structures in \mathcal{G} share the same domain. Then—having defined \mathcal{G} —we investigate the finite set S that contains, for *every* linear ordering $<_A$ of A and *every* structure in \mathcal{G} , the binary encoding of that structure with respect to $<_A$. We choose the string $s \in S$ with the least binary number. Let $G_s \in \mathcal{G}$ be a structure encoded by s . Using s , we choose the lexicographically smallest tuple \vec{c} from the winning relation of G_s that does not exhibit global losing symmetry.¹⁵ Thus, altogether, we obtain a renaming G_s of G together with the tuple $\vec{c} \in W_{G_s}$ that does not exhibit global losing symmetry.

Let c_j denote some coordinate of \vec{c} such that there exists a full renaming (β, π) from G_s to G that associates the player role number j with the player role number i . There may be several such coordinates j and several renamings (β, π) for j . Let D be the subset¹⁶ of the domain of G that contains exactly all choices c_i of player i in G that can be obtained by such renamings (β, π) and coordinates j . The desired protocol outputs D on the input (G, i) . \square

As discussed above, structural conventions may be quite artificial and arbitrary, and certainly cannot always be considered purely rational. It seems very difficult to separate, in a natural and commonly acceptable way, all purely rational principles from structural conventions.

¹⁵There can be more than one linear ordering $<_A$ leading to the string s , so there can be many ways to choose the tuple \vec{c} . This will not be a problem for us; we will see that it makes no difference which tuple \vec{c} we choose. Note that there will always be an isomorphism connecting different such tuples \vec{c} .

¹⁶It will become clear that D is independent of which tuple \vec{c} we chose.

6 Concluding remarks

We have proposed and studied a hierarchy of principles of rational coordination that can be applied in rational players’ reasoning about how to act in pure coordination scenarios. To make our study precise we have formalised such scenarios as WLC games and have compared the scope of applicability and strength of the various principles in terms of the classes of WLC games solvable by them. One major conclusion we draw is that the very questions of what is rational reasoning and what are rational choices in pure coordination games without preplay communication or conventions, appears to be very subtle and non-trivial. In particular, it seems very difficult to separate purely rational principles from others that can be employed in rational players’ reasoning when there is no purely rational solution. On the other hand, there is a precise technical description of the class of WLC games that are solvable when structural conventions are enabled.

A number of conceptual and technical issues arising from the present study remain open for further exploration.

- To begin with, in this paper we have focused on scenarios where players look for choices that *guarantee* winning if a suitable rational principle is followed. But it is very natural to ask how players should act in a game which seems not solvable by any purely rational principle¹⁷. If players cannot guarantee a win, it is natural to assume that they should at least try to maximize somehow their collective chances of winning, say, by considering protocols involving some *probability distribution* over their choices.
- Another natural extension of the present framework is to consider non-structural principles based on limited preplay communication and use of various types of non-structural conventions based on some additional features of the game representation, e.g., partial priority orders of players, colours of choices, etc. For a general exploration of WLC games with enriched game representation see [11].
- Studying pure dis-coordination games and games involving combinations of coordination and dis-coordination types of players are major potential directions for further work.

Finally, to provide contrast with the theoretical work presented in this paper, in the future we plan to run empirical experiments on people’s behaviour WLC games.

¹⁷We note that if players were ultimately interested *only* in *guaranteeing* a win in WLC-games, even the non-losing principle NL could be questioned. This is simply because, in structurally unsolvable games (such as $G(2(1 \times 1) + \overline{1 \times 1})$), winning is not guaranteed by any structural principles—regardless whether players follow NL or not. This demonstrates that rational players should, of course, not only be interested in guaranteeing a win when that is possible.

Acknowledgements

The work of Valentin Goranko was partly supported by a research grant 2015-04388 of the Swedish Research Council. The work of Antti Kuusisto was supported by the ERC grant 647289 “CODA” and a part of the work also by the Academy of Finland grants 438 874 and 209 365. The work of Raine Rönholm was partly supported by Jenny and Antti Wihuri Foundation. We thank the reviewers of this paper, as well as the reviewers of [13], [12], and [11], for helpful remarks and references.

References

- [1] Fausto Barbero. Cooperation in games and epistemic readings of independence-friendly sentences. *to appear in the Journal of Logic, Language and Information*, 2017.
- [2] Nicholas Bardsley, Judith Mehta, Chris Starmer, and Robert Sugden. Explaining focal points: Cognitive hierarchy theory versus team reasoning. *The Economic Journal*, 120:40–79, 2009.
- [3] B. Douglas Bernheim. Rationalizable strategic behavior. *Econometrica*, pages 1007–1028, 1984.
- [4] Cristina Bicchieri. *Rationality and Coordination*. Cambridge University Press, 1994.
- [5] Ken Binmore and Larry Samuelson. The Evolution of Focal Points. Economics Working Papers 0017, Institute for Advanced Study, School of Social Science, February 2002.
- [6] Heinz-Dieter Ebbinghaus, Jörg Flum, and Wolfgang Thomas. *Mathematical logic (2. ed.)*. Undergraduate texts in mathematics. Springer, 1994.
- [7] Drew Fudenberg and Jean Tirole. *Game theory*. MIT Press, 1991.
- [8] David Gauthier. Coordination. *Dialogue*, 14(02):195–221, 1975.
- [9] Michael R. Genesereth, Matthew L. Ginsberg, and Jeffrey S. Rosenschein. Cooperation without communication. In *Proc. of AAAI’86, vol. 1*, pages 51–57, 1986.
- [10] Margaret Gilbert. Rationality, coordination, and convention. *Synthese*, 84(1):1–21, 1990.
- [11] Valentin Goranko, Antti Kuusisto, and Raine Rönholm. Rational coordination in games with enriched representations. In *Proc. of EUMAS 2017*, pages 323–338, 2017.
- [12] Valentin Goranko, Antti Kuusisto, and Raine Rönholm. Rational coordination with no communication or conventions. In *Proc. of LORI VI*, volume 10455 of *LNCS*, pages 33–48. Springer, 2017.

- [13] Valentin Goranko, Antti Kuusisto, and Raine Rönholm. Rationality principles in pure coordination games. Extended abstract of a talk presented at the 5th International Workshop on Strategic Reasoning, July 2017, Liverpool, UK, 2017.
- [14] Peter Hawke. The logic of joint ability in two-player tacit games. *The Review of Symbolic Logic*, pages 1–28, 2017.
- [15] Maarten C.W. Janssen. Rationalizing focal points. *Theory and Decision*, 50(2):119–148, 2001.
- [16] D. Lewis. *Convention, A Philosophical Study*. Harvard University Press, 1969.
- [17] Kevin Leyton-Brown and Yoav Shoham. *Essentials of Game Theory: A Concise Multidisciplinary Introduction*. Morgan & Claypool Publishers, 2008.
- [18] Leonid Libkin. *Elements of Finite Model Theory*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2004.
- [19] Kerkko Luosto. Logical co-operation in multiplayer games. In *Programme and Abstracts, 2017 European Summer Meeting of the Association for Symbolic Logic, Logic Colloquium '17, Stockholm, Sweden, August 14–20*, page 156. ASL, 2017.
- [20] Judith Mehta, Chris Starmer, and Robert Sugden. Focal points in pure coordination games: An experimental investigation. *Theory and Decision*, 36(2):163–185, 1994.
- [21] Martin Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT Press, 1994.
- [22] David G. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica*, pages 1029–1050, 1984.
- [23] Hans Peters. *Game Theory: A Multi-Leveled Approach*. Springer, 2008.
- [24] Hofstadter D. R. Dilemmas for superrational thinkers, leading up to a luring lottery. *Scientific American*, 248(3), 1983.
- [25] Thomas C. Schelling. *The strategy of conflict*. Harvard University Press, 1960.
- [26] Robert Sugden. Spontaneous order. *J. of Economic Perspectives*, 3(4):85–97, 1989.
- [27] Robert Sugden. Rational Choice: A Survey of Contributions from Economics and Philosophy. *Economic Journal*, 101(407):751–785, July 1991.
- [28] Robert Sugden. A theory of focal points. *The Economic Journal*, pages 533–550, 1995.
- [29] Paul Syverson. *Logic, Convention, and Common Knowledge. A Conventionalist Account of Logic*. CSLI Lecture Notes 142. CSLI Publications, 2002.